

Optimal dual martingales, their analysis and application to new algorithms for Bermudan products^{1,2}

John Schoenmakers Junbo Huang Jianing Zhang

February 14, 2012

Abstract

In this paper we introduce and study the concept of optimal and surely optimal dual martingales in the context of dual valuation of Bermudan options, and outline the development of new algorithms in this context. We provide a characterization theorem, a theorem which gives conditions for a martingale to be surely optimal, and a stability theorem concerning martingales which are near to be surely optimal in a sense. Guided by these results we develop a framework of backward algorithms for constructing such a martingale. In turn this martingale may then be utilized for computing an upper bound of the Bermudan product. The methodology is purely dual in the sense that it doesn't require certain input approximations to the Snell envelope.

In an Itô-Lévy environment we outline a particular regression based backward algorithm which allows for computing dual upper bounds without nested Monte Carlo simulation. Moreover, as a by-product this algorithm also provides approximations to the continuation values of the product, which in turn determine a stopping policy. Hence, we may obtain lower bounds at the same time.

In a first numerical study we demonstrate the backward dual regression algorithm in a Wiener environment at well known benchmark examples. It turns out that the method is at least comparable to the one in Belomestny et. al. (2009) regarding accuracy, but regarding computational robustness there are even several advantages.

Keywords: Bermudan options, duality, Monte Carlo simulation, linear regression, surely optimal martingales, backward algorithm.

MSC: 62L15, 65C05, 91B28

¹This paper is an extended version of the previous preprint Schoenmakers and Huang (2010). Supported by DFG Research Center MATHEON "Mathematics for Key Technologies" in Berlin.

²Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, {schoenma,huang,zhang}@wias-berlin.de

1 Introduction

It is well-known that the evaluation of Bermudan callable derivatives comes down to solving an optimal stopping problem. For many callable exotic products, e.g. interest products, the underlying state space is high-dimensional however. As such these products are usually computationally expensive to solve with deterministic (PDE) methods and therefore simulation based (Monte Carlo) methods are called for. The first developments in this respect concentrated on the construction of a “good” exercise policy. We mention, among others, regression based methods by Carriere (1996), Longstaff and Schwartz (2001), and Tsitsiklis and Van Roy (2001), the stochastic mesh method of Broadie and Glasserman (2004), and quantization algorithms by Bally and Pages (2003). Especially for very high dimensions, Kolodko and Schoenmakers (2004) developed a policy improvement approach which can be effectively combined with Longstaff and Schwartz (2001) for example (see Bender et al. (2008) and Bender et al. (2006)).

As a common feature, the aforementioned simulation methods provide lower biased estimates for the Bermudan product under consideration. As a new breakthrough, Rogers (2002), and Haugh and Kogan (2004) introduced a dual approach, which comes down to minimizing over a set of martingales rather than maximizing over a family of stopping times. By its very nature the dual approach gives upper biased estimates for the Bermudan product and after its discovery several numerical algorithms for computing dual upper bounds have been proposed. Probably the most popular one is the method of Andersen and Broadie (2004), although this method requires nested Monte Carlo simulation (see also Kolodko and Schoenmakers (2004) and Schoenmakers (2005)). In a Wiener environment, Belomestny et. al. (2009) provides a fast generic method for computing dual upper bounds which avoids nested simulations. Further Brown et al. (2009) consider dual optimization via enlarging the information were an exercise decision may depend on. In this setting they also provide an example where a tight dual upper bound can be obtained by non-nested simulation.

The algorithms for computing dual upper bounds so far have in common that they start with some given “good enough” approximation of the Snell envelope and then construct the Doob martingale due to this approximation. In a recent paper Rogers (2010), points out how to construct a particular ‘good’ martingale via a sequence of martingales which are constant on an even bigger time interval. In this construction no input approximation to the Snell envelope is used. The methods proposed in this paper have some flavor of the method of Rogers (2010), in the sense that no approximation to the Snell envelope is involved either. In a recent paper Desai et. al. (2010) treat the dual problem by methods from convex optimization theory.

The structure of this paper is as follows. Starting with a short resume of well-known facts on Bermudan derivatives in Section 2, we analyze in Section 3 the almost sure property of the dual representation in detail. There we introduce the concept of a surely optimal martingale, which is loosely speaking, a martingale that minimizes the dual representation with a particular almost sure property.

In this respect we will point out that a martingale which minimizes the dual representation is not necessarily surely optimal, and on the other hand, a surely optimal martingale is generally not unique.

In Section 4 we present, as one of the main contributions of this paper, a characterization theorem for surely optimal martingales (Theorem 6). Moreover, we provide another result that guarantees that a martingale is surely optimal if it satisfies a certain measurability criterion (Theorem 10).

In applications of the algorithm of Andersen and Broadie (2004) one generally observes that the lower the variance of the upper bound estimator, i.e. the closer the corresponding martingale is to a surely optimal one, the sharper is the corresponding dual upper bound. Actually this observation was not well studied from a mathematical point of view so far. In Section 5 we study this phenomenon and, as a next main contribution, give an explanation of it by Theorem 12 and Corollary 15. In fact, the latter corollary may be considered a stability statement connected to Theorem 10.

Guided by the new theoretical insights we develop in Section 6 algorithms for constructing dual martingales that are based on minimization of the variance (respectively expected conditional variance) of corresponding dual representations and estimators. In this context we present in an Itô-Lévy environment a regression based backward procedure that constructs a dual martingale via minimizing backwardly in time the expected (conditional) variances of the dual estimators corresponding to the Snell envelope. We so obtain a martingale that allows for computing upper bounds without nested Monte Carlo (like in Belomestny et. al. (2009)). Moreover we obtain, as a by-product, estimations of continuation values. Thus, as a result, we end up with a procedure that computes upper bounds as well as lower bounds simultaneously via a non-nested simulation procedure. The procedure is quite easy to implement and may be considered as a valuable alternative to the non-nested method of Belomestny et. al. (2009), where a dual martingale is obtained by constructing a discretized Clark-Ocone derivative of some (input) approximation to the Snell envelope via regression. In particular, our new procedure only requires regression at each exercise date, in contrast to the procedure of Belomestny et. al. (2009) that requires regression at each time point of a sufficient refinement of the exercise grid.

In Section 7, we present a numerical study of our algorithm. We illustrate at two multi-dimensional benchmark products (one of which is also considered in Belomestny et. al. (2009)) a backward regression algorithm that, regarding accuracy and computational effort, produces upper bounds that show to be at least of the same quality as those in Belomestny et. al. (2009), and fast lower bounds that are overall better than in Belomestny et. al. (2009) moreover. In an Appendix, we provide standard results from Statistics which are used for several technical arguments in Section 6.1.

2 Bermudan derivatives and optimal stopping

Let $(Z_i : i = 0, 1, \dots, T)^1$ be a non-negative stochastic process in discrete time on a filtered probability space (Ω, \mathcal{F}, P) , adapted to a filtration $\mathbb{F} := (\mathcal{F}_i : 0 \leq i \leq T)$ which satisfies $E|Z_i| < \infty$, for $0 \leq i \leq T$. The measure P may be considered as a pricing measure and the process Z may be seen as a (discounted) cash-flow which an investor may exercise once in the time set $\{0, \dots, T\}$. Hence, she is faced with a Bermudan product. A well-known fact is that a fair price of such a derivative is given by the Snell envelope

$$Y_i^* = \sup_{\tau \in \{i, \dots, T\}} E_i Z_\tau, \quad 0 \leq i \leq T, \quad (1)$$

at time $i = 0$. In (1), τ denotes a stopping time, $E_i := E_{\mathcal{F}_i}$ denotes the conditional expectation with respect to the σ -algebra \mathcal{F}_i , and \sup (\inf) is to be understood as *essential supremum* (*essential infimum*) if it ranges over an uncountable family of random variables. Let us recall some well-known facts (e.g. see Neveu (1975)).

1. The Snell envelope Y^* of Z is the smallest super-martingale that dominates Z .
2. A family of optimal stopping times is given by

$$\tau_i^* = \inf\{j : j \geq i, \quad Z_j \geq Y_j^*\}, \quad 0 \leq i \leq T.$$

In particular,

$$Y_i^* = E_i Z_{\tau_i^*}, \quad 0 \leq i \leq T,$$

and the above family is the family of first optimal stopping times if several optimal stopping families exist.

The optimal stopping problem (1) has a natural interpretation from the point of view of the option holder: she seeks for an optimal exercise strategy which optimizes her expected payoff. On the other hand, the seller of the option rather seeks for the minimal cash amount (smallest supermartingale) he has to have at hand in any case the holder of the option exercises.

3 Duality and surely optimal martingales

We briefly recall the dual approach proposed by Rogers (2002) and, independently, Haugh and Kogan (2004). The dual approach is based on the following observation: for any martingale (M_j) with $M_0 = 0$ we have

$$Y_0^* = \sup_{\tau \in \{0, \dots, T\}} E_0 Z_\tau \leq \sup_{\tau \in \{0, \dots, T\}} E_0 (Z_\tau - M_\tau) \leq E_0 \max_{0 \leq j \leq T} (Z_j - M_j), \quad (2)$$

¹For notational convenience we have chosen for this stylized time set. The reader may reformulate all statements and results in this paper for a general discrete time set $\{T_0, T_1, \dots, T_J\}$ in a trivial way.

hence the right-hand side provides an upper bound for Y_0^* . Rogers (2002) and Haugh and Kogan (2004) showed that (2) holds with equality for the martingale part of the Doob decomposition of Y^* , i.e. $Y_j^* = Y_0^* + M_j^* - A_j^*$, where M^* is a martingale with $M_0^* = 0$, and A^* is predictable with $A_0^* = 0$. More precisely we have

$$M_j^* = \sum_{l=1}^j (Y_l^* - E_{l-1}Y_l^*), \quad A_j^* = \sum_{l=1}^j (Y_{l-1}^* - E_{l-1}Y_l^*), \quad (3)$$

from which we see A^* is non-decreasing due to Y^* being a supermartingale. In addition, they showed that

$$Y_0^* = \max_{0 \leq j \leq T} (Z_j - M_j^*) \quad \text{a.s.} \quad (4)$$

The next lemma, by Kolodko and Schoenmakers (2006), provides a somewhat more general class of supermartingales, which turns relation (2) into an equality such that moreover (4) holds.

Lemma 1 *Let S be a supermartingale with $S_0 = 0$. Assume that $Z_j - Y_0^* \leq S_j$, $1 \leq j \leq T$. It then holds that*

$$Y_0^* = \max_{0 \leq j \leq T} (Z_j - S_j) \quad \text{a.s.} \quad (5)$$

For the proof see Kolodko and Schoenmakers (2006).

Examples 2 Obviously, by taking for S the Doob martingale as constructed in (3), Lemma 1 applies. However, the Doob martingale is not the only one. For example, in the case $Z > 0$ a.s. we may also take

$$S_j = (N_j^* - 1)Y_0^*,$$

where N^* is the multiplicative Doob part of the Snell envelope. More precisely, $Y_j^* = Y_0^* N_j^* B_j^*$ for a martingale N^* with $N_0^* = 1$ and predictable B^* with $B_0^* = 1$. Hence

$$N_j^* = \prod_{l=1}^j \frac{Y_l^*}{E_{l-1}Y_l^*}, \quad B_j^* = \prod_{l=1}^j \frac{E_{l-1}Y_l^*}{Y_{l-1}^*}. \quad (6)$$

Indeed, since B^* is non-increasing due to Y^* being a supermartingale, we have

$$S_j = Y_0^* \left(\frac{Y_j^*}{Y_0^* B_j^*} - 1 \right) \geq Y_0^* \left(\frac{Y_j^*}{Y_0^*} - 1 \right) = Y_j^* - Y_0^* \geq Z_j - Y_0^*,$$

thus, Lemma 1 applies again.

The multiplicative Doob decomposition in (6) is used by Jamshidian (2007) for constructing a multiplicative dual representation. In a comparative study, Chen and Glasserman (2007) pointed out however, that from a numerical point of view additive dual algorithms perform better due to the nice almost sure property (4).

Remark 3 It is **not** true that for any martingale M which turns (2) into equality the almost sure statement (4) holds. As a simple counterexample, consider $T = 1$, $Z_0 = 0$, $Z_1 = 2$, $M_0 = 0$, and $M_1 = \pm 1$ each with probability $1/2$. Indeed, we see that $Y_0^* = 2 = E_0(2 - M_1) = E_0 \max(0, 2 - M_1)$, but, $Y_0^* \neq \max(0, 2 - M_1)$ a.s.

In order to have a unified dual representation for the Snell envelope Y_i^* at any i , it is convenient to drop the assumption that martingales start at zero. We then may restate the dual theorem as

$$Y_i^* = \inf_{M \in \mathcal{M}} E_i \max_{j \leq T} (Z_j - M_j + M_i) \quad (7)$$

$$= \max_{j \leq T} (Z_j - M_j^* + M_i^*) \quad \text{a.s.}, \quad (8)$$

for any i , $0 \leq i \leq T$, where \mathcal{M} is the set of all martingales and M^* is the Doob martingale part of Y^* .

In view of Remark 3 and Examples 2, a martingale for which the infimum (7) is attained must not necessarily satisfy an almost sure property such as (8), and, martingales which do satisfy such almost sure property are generally not unique. We hence propose the following concept of *surely optimal* martingales.

Definition 4 We say that a martingale M is **surely optimal** for the Snell envelope Y^* at a time i , $0 \leq i \leq T$, if it holds

$$Y_i^* = \max_{j \leq T} (Z_j - M_j + M_i) \quad \text{a.s.} \quad (9)$$

Remark 5 Obviously, the Doob martingale of Y^* is surely optimal at each i , $0 \leq i \leq T$, and any martingale M is trivially surely optimal at $i = T$. However, it is **not** true that sure optimality for some i with $i < T$ implies sure optimality at $i + 1$. As a counterexample let us consider $T = 2$, and $Z_0 = 4$, $Z_1 = 0$, $Z_2 = 2$. Take as martingale $M_0 = 0$, $M_1 = \pm 1$, each with probability $1/2$, and $M_2 = M_1 \pm 1$, each with probability $1/2$ conditional on M_1 . Then $\max_{0 \leq j \leq 2} (Z_j - M_j + M_0) = 4$ a.s. Since we have trivially $Y_0^* = 4$, M is surely optimal at $i = 0$. But, $\max_{1 \leq j \leq 2} (Z_j - M_j + M_1) = 2 - M_2 + M_1 \notin \mathcal{F}_1$, so M is not surely optimal for Y^* at $i = 1$.

4 Characterization of surely optimal martingales

In this section we give a characterization of martingales that are surely optimal for all $i = 0, \dots, T$.

Theorem 6 A martingale M with $M_0 = 0$ is surely optimal for $i = 0, \dots, T$, if and only if there exists a sequence of adapted random variables $(\zeta_i)_{0 \leq i \leq T}$, such that $E_{i-1} \zeta_i = 1$, and $\zeta_i \geq 0$ for all $0 < i \leq T$, and

$$M_i = M_i^* - A_i^* + \sum_{l=1}^i (A_l^* - A_{l-1}^*) \zeta_l, \quad (10)$$

where, respectively, M^* is the Doob martingale and A_i^* the predictable process of the Snell envelope Y^* as given in (3).

Proof. i) Let us assume that M is surely optimal as stated. Then by (9) it holds for any $0 < i \leq T$,

$$\begin{aligned} Y_{i-1}^* &= \max_{i-1 \leq j \leq T} (Z_j - M_j + M_{i-1}) \\ &= \max(Z_{i-1}, M_{i-1} - M_i + \max_{i \leq j \leq T} (Z_j - M_j + M_i)) \\ &= \max(Z_{i-1}, M_{i-1} - M_i + Y_i^*). \end{aligned} \quad (11)$$

Since $Z_{i-1} \leq Y_{i-1}^*$, and since $Z_{i-1} < Y_{i-1}^*$ implies $A_{i-1}^* = A_i^*$, we obtain from (11) and the Doob decomposition $Y_i^* = Y_0^* + M_i^* - A_i^*$

$$\begin{aligned} Y_{i-1}^* - Z_{i-1} &= (M_{i-1} - M_i + Y_i^* - Z_{i-1})^+ \\ &= (M_{i-1} - M_i + M_i^* - M_{i-1}^* - A_i^* + A_{i-1}^* + Y_{i-1}^* - Z_{i-1})^+ \\ &= 1_{Z_{i-1} < Y_{i-1}^*} (M_{i-1} - M_i + M_i^* - M_{i-1}^* + Y_{i-1}^* - Z_{i-1})^+ \\ &\quad + 1_{Z_{i-1} = Y_{i-1}^*} (M_{i-1} - M_i + M_i^* - M_{i-1}^* - A_i^* + A_{i-1}^*)^+. \end{aligned}$$

So we must have

$$\begin{aligned} 1_{Z_{i-1} < Y_{i-1}^*} (Y_{i-1}^* - Z_{i-1}) &= \\ 1_{Z_{i-1} < Y_{i-1}^*} (M_{i-1} - M_i + M_i^* - M_{i-1}^* + Y_{i-1}^* - Z_{i-1}), \quad \text{and} \\ 1_{Z_{i-1} = Y_{i-1}^*} (M_{i-1} - M_i + M_i^* - M_{i-1}^* - A_i^* + A_{i-1}^*)^+ &= 0, \end{aligned}$$

respectively. Hence we get

$$\begin{aligned} 1_{Z_{i-1} < Y_{i-1}^*} (M_{i-1} - M_i + M_i^* - M_{i-1}^*) &= 0, \quad \text{and} \\ 1_{Z_{i-1} = Y_{i-1}^*} (M_{i-1} - M_i + M_i^* - M_{i-1}^* - A_i^* + A_{i-1}^*) &= -1_{Z_{i-1} = Y_{i-1}^*} \mu_i, \end{aligned} \quad (12)$$

for some non-negative \mathcal{F}_i -measurable random variable μ_i . W.l.o.g. we assume that $\mu_i \equiv 0$ on the set $\{Z_{i-1} < Y_{i-1}^*\}$. By taking \mathcal{F}_{i-1} conditional expectations on both sides of (13), and using the martingale property of both M and M^* , and the predictability of A^* , it then follows that

$$E_{i-1} \mu_i = 1_{Z_{i-1} = Y_{i-1}^*} E_{i-1} \mu_i = 1_{Z_{i-1} = Y_{i-1}^*} (A_i^* - A_{i-1}^*). \quad (14)$$

In particular, since $\mu_i \geq 0$ almost surely, it follows from (14) that $\mu_i = 0$ on the set $\{A_i^* = A_{i-1}^*\}$ (in which $\{Z_{i-1} < Y_{i-1}^*\}$ is contained as a subset). We next define

$$\zeta_i := \begin{cases} (A_i^* - A_{i-1}^*)^{-1} \mu_i, & \text{if } A_i^* > A_{i-1}^*, \\ 1, & \text{else,} \end{cases} \quad (15)$$

and we see that we have a.s. $\mu_i = (A_i^* - A_{i-1}^*) \zeta_i$. By (14) we have (using the convention $0 \cdot \infty = 0$)

$$\begin{aligned} E_{i-1} \zeta_i &= 1_{A_i^* > A_{i-1}^*} E_{i-1} (A_i^* - A_{i-1}^*)^{-1} \mu_i + 1_{A_i^* = A_{i-1}^*} \\ &= 1_{A_i^* > A_{i-1}^*} 1_{Z_{i-1} = Y_{i-1}^*} + 1_{A_i^* > A_{i-1}^*} 1_{Z_{i-1} < Y_{i-1}^*} + 1_{A_i^* = A_{i-1}^*} = 1, \end{aligned}$$

since the middle term is trivially zero. We thus obtain from (12) and (13)

$$M_{i-1} - M_i + M_i^* - M_{i-1}^* - A_i^* + A_{i-1}^* = -(A_i^* - A_{i-1}^*) \zeta_i,$$

from which (10) follows.

ii) Conversely, if a martingale M satisfies (10), we have for any $0 \leq i \leq T$,

$$\begin{aligned} \max_{i \leq j \leq T} (Z_j - M_j + M_i) &= \max_{i \leq j \leq T} \left(Z_j - M_j^* + A_j^* - \sum_{l=1}^j (A_l^* - A_{l-1}^*) \zeta_l \right. \\ &\quad \left. + M_i^* - A_i^* + \sum_{l=1}^i (A_l^* - A_{l-1}^*) \zeta_l \right) \\ &= Y_i^* + \max_{i \leq j \leq T} \left(Z_j - Y_j^* - \sum_{l=i+1}^j (A_l^* - A_{l-1}^*) \zeta_l \right) \leq Y_i^*, \end{aligned}$$

and then by (7) the almost sure optimality follows. ■

By Theorem 6 we have immediately the following alternative characterization of almost sure martingales. It basically says that a martingale is surely optimal if the Snell envelope can be represented in a way that resembles the Doob decomposition but where the predictable process is replaced by a process which is in general only adapted.

Corollary 7 *A martingale M with $M_0 = 0$ is surely optimal for $i = 0, \dots, T$, if and only if there exists a non-decreasing adapted process N with $N_0 = 0$ such that²*

$$Y_i^* = Y_0^* + M_i - N_i.$$

Proof. If M is surely optimal as stated, we have by the “if” part of Theorem 6 (see (10)),

$$Y_i^* - Y_0^* - M_i = - \sum_{l=1}^i (A_l^* - A_{l-1}^*) \zeta_l = -N_i, \quad (16)$$

with N being adapted, non-decreasing and $N_0 = 0$. Conversely, if

$$Y_i^* = Y_0^* + M_i - N_i$$

for some martingale M , $M_0 = 0$, and non-decreasing adapted N , $N_0 = 0$, we consider for each i , $0 \leq i \leq T$,

$$\max_{i \leq j \leq T} (Z_j - M_j + M_i) = \max_{i \leq j \leq T} (Z_j - Y_j^* - N_j + Y_i^* + N_i) \leq Y_i^*,$$

and then apply (7) again. ■

We have the following remark.

²Note that N is not assumed to be predictable.

Remark 8 Let the martingale M with $M_0 = 0$ be surely optimal for $i = 0, \dots, T$. For the non-decreasing process N defined by (16) it holds that

$$Y_i^* - M_i + M_{i-1} - Z_{i-1} = Y_{i-1}^* - N_i + N_{i-1} - Z_{i-1} =: U_i,$$

and since by (16), $N_i - N_{i-1} = (A_i^* - A_{i-1}^*) \zeta_i$, we obtain from (11)

$$(U_i)^+ = Y_{i-1}^* - Z_{i-1} \quad \text{a.s.}$$

So, in particular we have that $(U_i)^+$ is \mathcal{F}_{i-1} -measurable while U_i itself is generally **not**, except for the case where $M = M^*$. A similar observation will encountered later on in (26).

From Theorem 6 it is clear that there exist infinitely many martingales which are surely optimal for all $i = 0, \dots, T$. In the following example we construct a one-parametric family of such martingales which includes the Doob martingale of the Snell envelope.

Example 9 Let us assume $Z > 0$ a.s. (if Z is strictly bounded from below by a constant $-K$, we may consider the equivalent stopping problem due to $Z + K$). Then $Y^* > 0$ a.s., and for any α , $0 \leq \alpha \leq 1$, we consider

$$\zeta_i := 1 - \alpha + \alpha \frac{Y_i^*}{E_{i-1} Y_i^*} = 1 - \alpha + \alpha \frac{N_l^*}{N_{l-1}^*},$$

where N^* is the martingale part of the multiplicative decomposition $Y_i^* = Y_0^* N_i^* B_i^*$ of the Snell envelope (see Examples 2). Obviously, it holds $E_{i-1} \zeta_i = 1$ and $\zeta_i \geq 0$, and hence, by Theorem 6 we obtain for every $0 \leq \alpha \leq 1$ a martingale

$$\begin{aligned} M_i &= M_i^* - A_i^* + \sum_{l=1}^i (A_l^* - A_{l-1}^*) \left(1 - \alpha + \alpha \frac{N_l^*}{N_{l-1}^*} \right) \\ &= M_i^* - \alpha A_i^* + \alpha \sum_{l=1}^i (A_l^* - A_{l-1}^*) \frac{N_l^*}{N_{l-1}^*}, \end{aligned}$$

which is surely optimal for $i = 0, \dots, T$. Thus, for $\alpha = 0$ (i.e. $\zeta_i \equiv 1$) we retrieve the standard Doob martingale of the Snell envelope, and for $\alpha = 1$ we obtain

$$\begin{aligned} M_i &= Y_i^* - Y_0^* + \sum_{l=1}^i (A_l^* - A_{l-1}^*) \frac{N_l^*}{N_{l-1}^*} \\ &= \sum_{l=1}^i \left(Y_l^* - Y_{l-1}^* + Y_{l-1}^* \left(1 - \frac{B_l^*}{B_{l-1}^*} \right) \frac{N_l^*}{N_{l-1}^*} \right) \\ &= Y_0^* \sum_{l=1}^i \left(N_l^* B_l^* - N_{l-1}^* B_{l-1}^* + B_{l-1}^* \left(1 - \frac{B_l^*}{B_{l-1}^*} \right) N_l^* \right) \\ &= Y_0^* \sum_{l=1}^i B_{l-1}^* (N_l^* - N_{l-1}^*). \end{aligned} \tag{17}$$

Note that this martingale differs from the martingale $Y_0^*(N_i^* - 1)$ from Example 2 (they would coincide after dropping the factors B_{i-1}^*). It is easy to show (using Theorem 6 again) that the latter martingale is in general only optimal at $i = 0$, while the martingale (17) is surely optimal for all $i = 0, \dots, T$, by construction.

The next theorem provides a key criterion for identifying surely optimal martingales.

Theorem 10 *Let Y^* be the Snell envelope of the cash-flow Z and let M be any martingale. Then, for any $i \in \{0, \dots, T\}$ it holds*

$$\max_{i \leq j \leq T} (Z_j - M_j + M_i) \in \mathcal{F}_i \implies \max_{i \leq j \leq T} (Z_j - M_j + M_i) = Y_i^*.$$

Proof. Let us suppose $\vartheta_i := \max_{i \leq j \leq T} (Z_j - M_j + M_i) \in \mathcal{F}_i$ and define the stopping time

$$\tau_i = \inf \{j \geq i : Z_j - M_j + M_i \geq \vartheta_i\}.$$

By the definition of ϑ_i we have $i \leq \tau_i \leq T$ almost surely. We thus have

$$Y_i^* \geq E_i Z_{\tau_i} \geq E_i (M_{\tau_i} - M_i + \vartheta_i) = \vartheta_i,$$

by Doob's optional sampling theorem and the fact that $\vartheta_i \in \mathcal{F}_i$. On the other hand we have $\vartheta_i = E_i \vartheta_i \geq Y_i^*$ due to (7). ■

Remark 11 *While in this paper we work in a discrete time setting, it is obvious that Theorem 10 can be proved in (almost) literally the same way for continuous time exercise as well.*

5 Stability of surely optimal martingales

In equivalent terms, Theorem 10 states that, if a martingale M is such that the conditional variance of

$$\vartheta_i(M) := \max_{i \leq j \leq T} (Z_j - M_j + M_i)$$

is zero for some $0 \leq i \leq T$, i.e.

$$\text{Var}_i \vartheta_i(M) := E_i (\vartheta_i(M) - E_i \vartheta_i(M))^2 = 0, \text{ a.s.,}$$

then $\vartheta_i(M) = Y_i^*$. Hence the martingale M is surely optimal at i . In this section we present a stability result for martingales M which are, loosely speaking, close to be surely optimal at some i , in the sense that $\text{Var}_i \vartheta_i(M)$ is small. More specifically, we provide mild conditions on a sequence of martingales $(M^{(n)})_{n \geq 1}$ which guarantee that the corresponding upper bounds converge to the Snell envelope in a sense, although the sequence of martingales $(M^{(n)})$ does not necessarily converge. We have the following result.

Theorem 12 Let $i \in \{0, \dots, T\}$. If $\text{Var}_i \vartheta_i^{(n)} \xrightarrow{P} 0$ for $n \rightarrow \infty$, where $\vartheta_i^{(n)} := \vartheta_i(M^{(n)})$, and if in addition the sequence of martingales $(M_i^{(n)})_{n \geq 1}$ is uniformly integrable, then it holds

$$E_i \vartheta_i^{(n)} \xrightarrow{L^1} Y_i^*.$$

Proof. Fix an $i \in \{0, \dots, T\}$ and suppose that the assumptions of the theorem are satisfied. Now take an $\epsilon > 0$. By introducing an auxiliary time $\partial > T$ and setting $Z_\partial = 0$ we next define the stopping time

$$\tau_i^{(n)} = \inf \left\{ j \geq i : Z_j - M_j^{(n)} + M_i^{(n)} \geq E_i \vartheta_i^{(n)} - \epsilon \right\} \wedge \partial.$$

We thus have with $M_\partial^{(n)} := M_T^{(n)}$, $n \geq 1$,

$$\begin{aligned} Y_i^* &\geq E_i Z_{\tau_i^{(n)}} = E_i Z_{\tau_i^{(n)}} 1_{\tau_i^{(n)} < \partial} \geq E_i \left(M_{\tau_i^{(n)}}^{(n)} - M_i^{(n)} + E_i \vartheta_i^{(n)} - \epsilon \right) 1_{\tau_i^{(n)} < \partial} \\ &= E_i \left(M_{\tau_i^{(n)}}^{(n)} - M_i^{(n)} + E_i \vartheta_i^{(n)} - \epsilon \right) - E_i \left(M_T^{(n)} - M_i^{(n)} + E_i \vartheta_i^{(n)} - \epsilon \right) 1_{\tau_i^{(n)} = \partial} \\ &= E_i \vartheta_i^{(n)} - \epsilon - E_i \left(M_T^{(n)} - M_i^{(n)} + E_i \vartheta_i^{(n)} - \epsilon \right) 1_{\tau_i^{(n)} = \partial} \quad \text{a.s.}, \end{aligned}$$

hence

$$\begin{aligned} E_i \vartheta_i^{(n)} &\leq Y_i^* + \epsilon + E_i \left| M_T^{(n)} - M_i^{(n)} + E_i \vartheta_i^{(n)} - \epsilon \right| 1_{\tau_i^{(n)} = \partial} \\ &=: Y_i^* + \epsilon + E_i U_i^{(n)} 1_{\tau_i^{(n)} = \partial} \quad \text{a.s.} \end{aligned} \tag{18}$$

Now it is easy to see that the family of random variables $(U_i^{(n)})_{n \geq 1}$ is uniformly integrable too. We so may take $K_\epsilon > 0$ such that

$$\sup_{n \geq 0} E U_i^{(n)} 1_{U_i^{(n)} > K_\epsilon} \leq \epsilon.$$

Further observe that due to a conditional version of Chebyshev's inequality,

$$0 \leq E_i 1_{\tau_i^{(n)} = \partial} = E_i 1_{\vartheta_i^{(n)} < E_i \vartheta_i^{(n)} - \epsilon} \leq \frac{\text{Var}_i \vartheta_i^{(n)}}{\epsilon^2} \xrightarrow{P} 0.$$

Since the family $(E_i 1_{\tau_i^{(n)} = \partial})_{n \geq 0}$ is bounded by 1, it is uniformly integrable. Hence, it follows that

$$E_i 1_{\tau_i^{(n)} = \partial} \xrightarrow{L^1} 0. \tag{19}$$

We thus have

$$\begin{aligned} E U_i^{(n)} 1_{\tau_i^{(n)} = \partial} &= E U_i^{(n)} 1_{U_i^{(n)} > K_\epsilon} 1_{\tau_i^{(n)} = \partial} + E U_i^{(n)} 1_{U_i^{(n)} \leq K_\epsilon} 1_{\tau_i^{(n)} = \partial} \\ &\leq \epsilon + K_\epsilon E 1_{U_i^{(n)} \leq K_\epsilon} 1_{\tau_i^{(n)} = \partial} \leq \epsilon + K_\epsilon E E_i 1_{\tau_i^{(n)} = \partial} < 2\epsilon \end{aligned}$$

for $n > N_{\epsilon, K_\epsilon}$ by (19). So for $n > N_{\epsilon, K_\epsilon}$, we derive from (18)

$$E\vartheta_i^{(n)} \leq EY_i + \epsilon + EU_i^{(n)}1_{\tau_i^{(n)}=\partial} \leq EY_i^* + 3\epsilon.$$

Thus,

$$\overline{\lim}_{n \rightarrow \infty} E\vartheta_i^{(n)} \leq EY_i^* + 3\epsilon$$

Since $\epsilon > 0$ was arbitrary,

$$\overline{\lim}_{n \rightarrow \infty} E\vartheta_i^{(n)} \leq EY_i^*.$$

On the other hand, due to (7) we have $E_i\vartheta_i^{(n)} \geq Y_i^*$ a.s. for all n , so

$$0 \leq \overline{\lim}_{n \rightarrow \infty} E \left| E_i\vartheta_i^{(n)} - Y_i^* \right| = \overline{\lim}_{n \rightarrow \infty} \left(E\vartheta_i^{(n)} - EY_i^* \right) \leq 0,$$

which finally proves $E_i\vartheta_i^{(n)} \xrightarrow{L_1} Y_i^*$. ■

Remark 13 Like Theorem 10, Theorem 12 can be formulated in a continuous time setting as well with (almost) literally the same proof.

The following simple example illustrates that Theorem 12 would not be true when the uniform integrability condition is dropped.

Example 14 Take $T = 1$, $Z_0 = Z_1 = 0$, $M_0^{(n)} = 0$, $M_1^{(n)} =: -\xi_n$ with $E_0\xi_n = 0$, $n = 1, 2, \dots$. Then obviously $Y_0^* = 0$, and we have

$$\vartheta_0^{(n)} = \max(Z_0 - M_0^{(n)}, Z_1 - M_1^{(n)}) = \max(0, \xi^{(n)}) = \xi_+^{(n)}.$$

Now take

$$\xi^{(n)} = \begin{cases} 1 & \text{with Prob. } \frac{n-1}{n} \\ 1-n & \text{with Prob. } \frac{1}{n} \end{cases}$$

(hence $E_0\xi^{(n)} = 0$). Then, for $n \rightarrow \infty$ we have $\text{Var}_0\vartheta_0^{(n)} = E_0(\xi_+^{(n)})^2 - (E_0\xi_+^{(n)})^2 = \frac{n-1}{n} - \left(\frac{n-1}{n}\right)^2 = \frac{n-1}{n^2} \rightarrow 0$, whereas $E_0\vartheta_0^{(n)} = E_0\xi_+^{(n)} = \frac{n-1}{n} \rightarrow 1$. Clearly, for each $K > 1$, $E_0 \left| M_1^{(n)} \right| 1_{\{|M_1^{(n)}| > K\}} \geq \frac{n-1}{n} 1_{\{n-1 > K\}} \rightarrow 1$ as $n \rightarrow \infty$, hence the $(M_1^{(n)})$ are not uniformly integrable.

In view of the next Corollary, Theorem 12 may be considered as a stability theorem related to Theorem 10.

Corollary 15 Let \mathcal{M}^{UI} be a set of uniformly integrable martingales. Then for any $i \in \{0, \dots, T\}$ it holds: For every $\epsilon > 0$ there exist a $\delta > 0$ such that

$$[M \in \mathcal{M}^{UI} \quad \text{and} \quad E \text{Var}_i \vartheta_i(M) < \delta] \implies 0 \leq E \vartheta_i(M) - Y_i^* < \epsilon.$$

Proof. Suppose the statement is not true for some i . Then there exists an $\epsilon_0 > 0$ such that for all $n \in \mathbb{N}$ there exists a martingale $M^{(n)} \in \mathcal{M}^{UI}$, for which $E \text{Var}_i \vartheta_i(M^{(n)}) < 1/n$ and $E \vartheta_i(M^{(n)}) - Y_i^* \geq \epsilon_0$. Since convergence in L_1 implies convergence in probability along a subsequence (indexed again by n) we thus have $\text{Var}_i \vartheta_i(M^{(n)}) \xrightarrow{P} 0$, and $E |\vartheta_i(M^{(n)}) - Y_i^*| \geq \epsilon_0$ along this subsequence. This contradicts Theorem 12. ■

Remark 16 *Theorem 12 is important in practical situations, for instance, for (possibly high dimensional) underlyings of jump-diffusion type in a Lévy-Itô setup. In this environment we may consider the following class of uniformly integrable martingales.*

Let W be an m -dimensional Brownian motion and let N denote a Poisson random measure, independent of W , with (deterministic) compensator measure $\nu(s, du)ds$ such that

$$\int_0^t \int_{\mathbb{R}^q} (u^2 \wedge 1) \nu(s, du) ds < \infty, \quad 0 \leq t \leq T.$$

Let $(\mathcal{F}_t)_{0 \leq t \leq T}$ be the filtration generated by W and N , augmented by null sets. Now let X be a D -dimensional Markov process, adapted to (\mathcal{F}_t) , and consider the mappings $c : [0, T] \times \mathbb{R}^D \rightarrow \mathbb{R}_{\geq 0}$ and $d : [0, T] \times \mathbb{R}^D \times \mathbb{R}^q \rightarrow \mathbb{R}_{\geq 0}$ satisfying

$$E \int_0^T |c(s, X_s)|^2 ds < \infty, \quad E \int_0^T \int_{\mathbb{R}^q} |d(s, X_s, u)|^2 \nu(s, du) ds < \infty. \quad (20)$$

We define the class of uniformly integrable martingales, \mathcal{M}^{UI} , as the set of all martingales M satisfying

$$\begin{aligned} M_t &= M_0 + M_t^c + M_t^d \\ &= M_0 + \int_0^t \varphi^c(s, X_s) dW_s + \int_0^t \int_{\mathbb{R}^q} \varphi^d(s, X_s, u) \tilde{N}(ds, du), \end{aligned}$$

where φ^c and φ^d satisfy

$$|\varphi^c| \leq c, \quad |\varphi^d| \leq d,$$

and $\tilde{N} = N - \nu$ is the compensated Poisson measure. Note that M is indeed a martingale and that the expected quadratic variation of M is given by

$$\begin{aligned} E [M, M]_t &= E \int_0^t |\varphi^c(s, X_s)|^2 ds + E \int_0^t \int_{\mathbb{R}^q} |\varphi^d(s, X_s, u)|^2 \nu(s, du) ds \\ &\leq E \int_0^t |c(s, X_s)|^2 ds + E \int_0^t \int_{\mathbb{R}^q} |d(s, X_s, u)|^2 \nu(s, du) ds. \end{aligned}$$

We then have for every $t \in [0, T]$,

$$\sup_{M \in \mathcal{M}^{UI}} E |M_t|^2 \leq \sup_{M \in \mathcal{M}^{UI}} E \sup_{0 \leq t \leq T} |M_t|^2 \leq \sup_{M \in \mathcal{M}^{UI}} CE[M, M]_T < \infty,$$

where the second estimation results from the Burkholder-Davis-Gundy inequality and the third estimation follows from (20). Finally, an application of the de la Vallée Poussin criterion yields that \mathcal{M}^{UI} is indeed a family of uniformly integrable martingales.

6 New dual algorithms for pricing of Bermudan derivatives

In this section we consider the design of a new dual algorithm for solving multiple stopping problems, hence pricing Bermudan products, which are based on the theoretical insights from Theorem 10, Theorem 12, and Corollary 15. In the following, we provide an assessment of the merits of variance minimizing dual algorithms based on these results.

6.1 Merits of variance minimizing dual algorithms

Let Q be some index set and $\mathcal{M} = \{M^q : q \in Q\}$ be a set of uniformly integrable martingales such that \mathcal{M} contains a martingale M^{q^*} which is surely optimal at $i = 0$. Suppose that for any $q \in Q$ we have N samples of $\vartheta_0(M^{q,(n)})$, $n = 1, \dots, N$. Based on these samples we may estimate $\text{Var}_0 \vartheta_0(M^q) = \text{Var} \vartheta_0(M^q)$ as usual by

$$\begin{aligned} \text{Var}^{(N)} \vartheta_0(M^q) &:= \frac{1}{N-1} \sum_{n=1}^N \left(\vartheta_0(M^{q,(n)}) - \overline{\vartheta_0(M^q)_N} \right)^2, \quad \text{with} \\ \overline{\vartheta_0(M^q)_N} &:= \frac{1}{N} \sum_{n=1}^N \vartheta_0(M^{q,(n)}). \end{aligned} \quad (21)$$

So, in principle, only two realizations ($N = 2$) would be enough to identify a q^* such that

$$0 = \text{Var} \vartheta_0(M^{q^*}) = \text{Var}^{(2)} \vartheta_0(M^{q^*}) = \min_{q \in Q} \sum_{n=1}^2 \left(\vartheta_0(M^{q,(n)}) - \overline{\vartheta_0(M^q)_N} \right)^2,$$

and then obtain $Y_0^* = \vartheta_0(M^{q^*}) = \vartheta_0(M^{q^*,(1)})$. Due to this stylized argumentation we may expect that in a case where although the set \mathcal{M} doesn't contain a martingale that is surely optimal at $i = 0$ but at least one martingale M^q such that $\text{Var} \vartheta_0(M^q)$ is “small enough”, we only need a relatively small sample size N to identify this martingale, leading to a tight upper bound $Y_0^{up} := E\vartheta_0(M^q)$. In the following, we formalize this idea by giving precise estimates for the variance estimators in terms of the family of uniformly integrable martingales.

Suppose we want to obtain an upper bound which is bounded from above by $Y_0^* + \epsilon$ for some given $\epsilon > 0$. Consider a family of uniformly integrable

martingales $\mathcal{M} = \{M^q : q \in Q\}$ which is “rich enough” in the sense that there exists a $\delta > 0$ according to Corollary 15, such that

$$\{M \in \mathcal{M} : \text{Var } \vartheta_0(M) < \kappa\delta\} \neq \emptyset,$$

for some $0 < \kappa < 1$. Then, in particular, we have $\inf_{q \in Q} \text{Var } \vartheta_0(M^q) < \kappa\delta$ which implies that for any $q \in Q$ we have $E\text{Var } \vartheta_0(M^q) = \text{Var } \vartheta_0(M^q) < \delta$. Now Corollary 15 again yields $0 \leq E\vartheta_0(M^q) - Y_0^* < \epsilon$. For a given set of realizations $\vartheta_0(M^{q,(n)})$, $n = 1, \dots, N$, $q \in Q$, we now may try to find an optimal $q_N^\circ \in Q$ arising as the solution of the minimization problem

$$\inf_{q \in Q} \text{Var}^{(N)} \vartheta_0(M^q) = \text{Var}^{(N)} \vartheta_0(M^{q_N^\circ}). \quad (22)$$

For convenience, we assume that such a q_N° does exist. Furthermore, we assume the existence of $q^\circ \in Q$ which satisfies

$$\inf_{q \in Q} \text{Var } \vartheta_0(M^q) = \text{Var } \vartheta_0(M^{q^\circ}).$$

Let $0 \leq \alpha \ll 1$ be a small threshold probability. Then one can show under mild conditions on the family of random variables $\{\vartheta_0(M^q) : q \in Q\}$, that for some constant $C > 0$ and quantile coefficient $c_\alpha > 0$ (only depending on α) we have with probability larger than $1 - \alpha$,

$$\begin{aligned} \text{Var}^{(N)} \vartheta_0(M^q) &\leq \text{Var } \vartheta_0(M^q) \left(1 + c_\alpha \sqrt{\frac{C}{N}}\right), \\ \left(1 - c_\alpha \sqrt{\frac{C}{N}}\right) \text{Var } \vartheta_0(M^q) &\leq \text{Var}^{(N)} \vartheta_0(M^q), \quad \text{for } q \in \{q^\circ, q_N^\circ\} \end{aligned} \quad (23)$$

(see Appendix for details). Thus, with probability larger than $1 - \alpha$,

$$\begin{aligned} \left(1 - c_\alpha \sqrt{\frac{C}{N}}\right) \text{Var } \vartheta_0(M^{q_N^\circ}) &\leq \text{Var}^N \vartheta_0(M^{q_N^\circ}) \leq \text{Var}^N \vartheta_0(M^{q^\circ}) \\ &\leq \text{Var } \vartheta_0(M^{q^\circ}) \left(1 + c_\alpha \sqrt{\frac{C}{N}}\right), \end{aligned}$$

which yields

$$\text{Var } \vartheta_0(M^{q_N^\circ}) \leq \kappa\delta \frac{1 + c_\alpha \sqrt{\frac{C}{N}}}{1 - c_\alpha \sqrt{\frac{C}{N}}}.$$

This implies that for

$$N = \inf \left\{ n : \kappa \frac{1 + c_\alpha \sqrt{\frac{C}{n}}}{1 - c_\alpha \sqrt{\frac{C}{n}}} < 1 \right\} \quad (24)$$

we have $P [\text{Var } \vartheta_0(M^{q_N^\circ}) < \delta] > 1 - \alpha$. Thus, by Corollary 15, the dual upper bound due to the martingale $M^{q_N^\circ}$ identified by (22) falls below $Y_0^* + \epsilon$ with probability larger than $1 - \alpha$. As a main feature, equation (24) demonstrates that the smaller κ , the fewer samples N we may choose for the identification of q_N° .

The above argumentation suggests to minimize the estimated variance of the dual estimator over a parametric set of martingales using a relatively small sample size N . However, as the parametric set of martingales \mathcal{M} needs to be “rich enough”, in practice there may be many parameters involved, which in turn may lead to a non-convex minimization problem with many local minima. As a remedy to this problem, rather than directly minimizing the variance of the dual estimator at time zero, we propose to minimize backwardly the expected conditional variances $E \text{Var}_i \vartheta_i(M^q)$ over $q \in Q$, starting from $i = T$ (where the conditional variance is trivially zero) down to $i = 0$, using a simple but effective recursive relationship between $\vartheta_i(M^q)$ and $\vartheta_{i+1}(M^q)$ as explained in the next subsection. For this backward minimization procedure the arguments above apply as well and moreover, as we will see, it opens the possibility for linear regression, hence also the possibility for fast numerical implementations.

6.2 Backward dual variance minimization

Motivated by Section 6.1 we now develop a backward recursive simulation based algorithm for the construction of a dual martingale M that yields tight upper bounds. In view of a such a Monte Carlo approach, we assume a Markov setting generated by some underlying Markov process $X := (X_t)_{0 \leq t \leq T}$, and a cash-flow of the form $Z_j := Z_j(X_j) := Z(j, X_j)$. First we describe the algorithm in a pseudo language which involves terms such as conditional expectations and conditional variances. Then, we spell out an implementable Monte Carlo algorithm where these expressions are replaced by their empirical counterparts.

To start out on a pseudo algorithmic level we construct a martingale M backwardly in a recursive way by establishing that from $i = T$ down to $i = 0$ the expected conditional variances $E \text{Var}_i \vartheta_i(M)$ are “as small as possible” in a sense that we will describe. The martingale M is such that for $j > i$, any increment

$$M_j - M_i \text{ is measurable with respect to } \Delta \mathcal{F}_{i,j} := \sigma\{X_s : i \leq s \leq j\}. \quad (25)$$

It is easy to see that the Doob martingale of the Snell envelope meets this measurability property, however, in general Theorem 6 yields that there may exist many other surely optimal martingales satisfying this property.

A corner stone of the whole procedure is the following recursion that holds

for any martingale M and any $i < T$,

$$\begin{aligned}\vartheta_i(M) &= \max \left(Z_i, \max_{i+1 \leq j \leq T} (Z_j - M_j + M_i) \right) \\ &= \max (Z_i, \vartheta_{i+1}(M) + M_i - M_{i+1}) \\ &= Z_i + (\vartheta_{i+1}(M) + M_i - M_{i+1} - Z_i)^+. \end{aligned} \quad (26)$$

Obviously, at every $i = 0, \dots, T$, $\vartheta_i(M)$ only depends on $(M_j - M_i)_{i \leq j \leq T}$, and at the starting time $i = T$ we initially have $\vartheta_T(M) = Z_T$ which trivially satisfies $E \text{Var}_T(\vartheta_T(M)) = 0$. Note that if M were already surely optimal, i.e. $\vartheta(M)$ were already equal to Y^* , then Remark 8 would imply that for $U_i := \vartheta_{i+1}(M) + M_i - M_{i+1} - Z_i$, $(U_i)^+ = \vartheta_i(M) - Z_i = Y_i^* - Z_i$ is already \mathcal{F}_i -measurable.

Now the essential idea is comprised in the following backward induction: Assume that for $i+1 \leq T$ we have constructed the increments $(M_j - M_{i+1})_{i+1 \leq j \leq T}$ and $\vartheta_{i+1}(M)$. Now the task is to find a random variable ξ_{i+1} such that

$$\xi_{i+1} \text{ is } \Delta\mathcal{F}_{i,i+1}\text{-measurable,} \quad E_i \xi_{i+1} = 0, \quad (27)$$

that solves the following minimization problem

$$\begin{aligned}\xi_{i+1} &:= \arg \min_{\xi \in \Delta\mathcal{F}_{i,i+1}, E_i \xi = 0} E \text{Var}_i \vartheta_i(M(\xi)) \\ &= \arg \min_{\xi \in \Delta\mathcal{F}_{i,i+1}, E_i \xi = 0} E \text{Var}_i (\vartheta_{i+1}(M) - \xi - Z_i)^+. \end{aligned} \quad (28)$$

Intuitively, ξ_{i+1} represents the optimal martingale increment and thus, we put $M_j(\xi_{i+1}) - M_i(\xi_{i+1}) := M_j - M_{i+1} + \xi_{i+1}$ for $j \geq i+1$. By construction, the random variable ξ_{i+1} satisfies (27), therefore, we obtain a set of martingale increments $(M_j(\xi_{i+1}) - M_i(\xi_{i+1}))_{i \leq j \leq T}$, which has now been extended from $j = i+1$ to $j = i$ and which satisfies for $j \geq i+1$,

$$\begin{aligned}M_j(\xi_{i+1}) - M_{i+1}(\xi_{i+1}) &= M_j(\xi_{i+1}) - M_i(\xi_{i+1}) + M_i(\xi_{i+1}) - M_{i+1}(\xi_{i+1}) \\ &= M_j - M_{i+1}\end{aligned}$$

and by construction also the measurability requirement (25). Now we extend the increments $(M_j - M_{i+1})_{i+1 \leq j \leq T}$ from $j = i+1$ to $j = i$ by setting

$$(M_j - M_i)_{i \leq j \leq T} = (M_j(\xi_{i+1}) - M_i(\xi_{i+1}))_{i \leq j \leq T}$$

Finally we put

$$\vartheta_i(M) = Z_i + (\vartheta_{i+1}(M) - \xi_{i+1} - Z_i)^+.$$

After carrying out these steps backwardly from $i = T$ down to $i = 0$ we end up with a family of martingale increments $(M_j - M_0)_{0 \leq j \leq T}$, hence a martingale $(M_j)_{0 \leq j \leq T}$, as $M_0 = 0$ without loss of generality. This martingale will be subsequently used to compute a dual upper bound for Y_0^* via

$$Y_0^{up} = E \max_{0 \leq j \leq T} (Z_j - M_j).$$

The key step in the above procedure is to find a solution to minimization problem (28): Suppose that the martingale increments satisfying (27) for some fixed i may be parametrized as $\xi_{i+1}(\beta)$ where β is some generic parameter. Based on a set of simulated trajectories of X one may then estimate for some β (which we specify in more details below) the conditional variance

$$E \text{Var}_i (\vartheta_{i+1}(M) - \xi_{i+1}(\beta) - Z_i)^+ =: E \text{Var}_i U_i^+(\beta)$$

by using e.g. kernel estimators (e.g. see Liero (1989)), and next minimize with respect to β . In particular when the dimension of the parameter space is very small (typically one-dimensional) this may lead to a feasible Monte Carlo procedure. However, if the set of martingale increments $\xi_{i+1}(\beta)$ is “rich enough” and is moreover linearly structured in β , that is

$$\xi_{i+1}(\beta) = \sum_{k=1}^K \beta_k \mathbf{m}_{i+1}^{(k)},$$

where $\beta = (\beta_1, \dots, \beta_K) \in \mathbb{R}^K$ and the random variables $\mathbf{m}_{i+1}^{(k)}, k = 1, \dots, K$, satisfy (27) for $K \geq 1$ sufficiently large, it is in general more effective to solve the dominating problem

$$\begin{aligned} \arg \min_{\beta \in \mathbb{R}^K} E \text{Var}_i U_i(\beta) &:= \arg \min_{\beta \in \mathbb{R}^K} E \text{Var}_i (\vartheta_{i+1}(M) - \xi_{i+1}(\beta) - Z_i) \\ &= \arg \min_{\beta \in \mathbb{R}^K} E \text{Var}_i \left(\vartheta_{i+1}(M) - \sum_{k=1}^K \beta_k \mathbf{m}_{i+1}^{(k)} \right). \end{aligned} \quad (29)$$

The reason is twofold. On the one hand, if we succeed to find $\beta^\circ \in \mathbb{R}^K$ such that $E \text{Var}_i U_i(\beta^\circ)$ is sufficiently small (if it were zero, we would have arrived at an surely optimal martingale increment), then since

$$\arg \min_{\beta \in \mathbb{R}^K} E \text{Var}_{X_{T_i}} U_i^+(\beta) \leq E \text{Var}_i U_i^+(\beta^\circ) \leq E \text{Var}_i U_i(\beta^\circ),$$

$E \text{Var}_i U_i^+(\beta^\circ)$ is generally even closer to zero and so β° can be considered a good approximation to (28) as well. On the other hand, most importantly, problem (29) can be treated as a linear regression problem,

$$[\beta^\circ, \gamma^\circ] = \arg \min_{\beta \in \mathbb{R}^K, \gamma \in \mathbb{R}^{K_1}} E \left| \vartheta_{i+1}(M) - \sum_{k=1}^K \beta_k \mathbf{m}_{i+1}^{(k)} - \sum_{k=1}^{K_1} \gamma_k \psi_k(i, X_i) \right|^2, \quad (30)$$

which employs an additional set of basis functions $\psi_k(t, x), k = 1, \dots, K_1$. To see

this, note that (30) is equivalent with

$$\begin{aligned}
[\beta^\circ, \gamma^\circ] &= \arg \min_{\beta \in \mathbb{R}^K, \gamma \in \mathbb{R}^{K_1}} EE_i \left(\vartheta_{i+1}(M) - \sum_{k=1}^K \beta_k \mathbf{m}_{i+1}^{(k)} - E_i \vartheta_{i+1}(M) \right. \\
&\quad \left. + E_i \vartheta_{i+1}(M) - \sum_{k=1}^{K_1} \gamma_k \psi_k(i, X_i) \right)^2 \\
&= \arg \min_{\beta \in \mathbb{R}^K, \gamma \in \mathbb{R}^{K_1}} \left\{ E \text{Var}_i \left(\vartheta_{i+1}(M) - \sum_{k=1}^K \beta_k \mathbf{m}_{i+1}^{(k)} \right) \right. \\
&\quad \left. + E \left(E_i \vartheta_{i+1}(M) - \sum_{k=1}^{K_1} \gamma_k \psi_k(i, X_i) \right)^2 \right\},
\end{aligned}$$

hence β° satisfies (29), and moreover for γ° it holds

$$\gamma^\circ = \arg \min_{\gamma \in \mathbb{R}^{K_1}} E \left(E_i \vartheta_{i+1}(M) - \sum_{k=1}^{K_1} \gamma_k \psi_k(i, X_i) \right)^2. \quad (31)$$

Further, the regression procedure (30) delivers as by-product

$$C_i(x) := \sum_{k=1}^{K_1} \gamma_k^\circ \psi_k(i, x),$$

an approximate continuation function that may be used afterwards to define a stopping rule and to simulate a corresponding lower biased estimation of Y_0^* .

Remark 17 (i) *In virtually all practical applications we are in a setting as described in Remark 16. In this environment we may model ξ_{i+1} as linear combinations of the form*

$$\begin{aligned}
\xi_{i+1}(\beta) &:= \sum_{k=1}^{N_1} \beta_k^c \int_{T_i}^{T_{i+1}} \varphi_k^c(s, X_s) dW_s \\
&\quad + \sum_{k=1}^{N_2} \beta_k^d \int_{T_i}^{T_{i+1}} \varphi_k^d(s, X_s, u) d\tilde{N}(ds, du),
\end{aligned} \quad (32)$$

where $N_1 + N_2 = K$ and $\varphi_k^c(s, x)$ and $\varphi_k^d(s, x, u)$ are suitable sets of basis functions satisfying the conditions in Remark 16. In this setting, we have

$$\mathbf{m}_{i+1}^{(k)} = \int_{T_i}^{T_{i+1}} \varphi_k^c(s, X_s) dW_s + \int_{T_i}^{T_{i+1}} \varphi_k^d(s, X_s, u) d\tilde{N}(ds, du)$$

and $\beta = (\beta_1^c, \dots, \beta_{N_1}^c, \beta_1^d, \dots, \beta_{N_2}^d) \in \mathbb{R}^K$.

As an alternative, we may also take

$$\xi_{i+1}(\beta) := \sum_{k=1}^K \beta_k \left(B_{i+1}^{(k)} - B_i^{(k)} \right) \quad (33)$$

for an arbitrary given set of discounted tradables $\left(B_j^{(k)} \right)_{0 \leq j \leq T}$ where the $B_j^{(k)} = B^{(k)}(j, X_j)$ are provided by some specific problem under consideration. For example it may happen that discounted European options are available in closed form. In any case, (32) and (33) satisfy the requirements (27) for any vector parameter $\beta \in \mathbb{R}^K$.

(ii) Suppose that the system of basis martingale increments and basis functions in the regression based minimization (30) is sufficiently “rich” that there even exist $\beta^{\circ\circ}$ and $\gamma^{\circ\circ}$ such that

$$\vartheta_{i+1}(M) - \sum_{k=1}^K \beta_k^{\circ\circ} \mathbf{m}_{i+1}^{(k)} - \sum_{k=1}^{K_1} \gamma_k^{\circ\circ} \psi_k(i, X_i) = 0 \quad a.s.$$

then one would need only one trajectory for X to identify $\beta^{\circ\circ}$ and $\gamma^{\circ\circ}$ via (30). This is a similar situation as discussed in Section 6.1: In practice when the system (32) is rich enough, a relatively low sample size will be sufficient to solve (30) effectively. This phenomenon will be confirmed by our experiments in Section 7.

Description of the Monte Carlo algorithm

Let us now spell out the empirical, implementable counterpart of the procedure described above. Based on a set of trajectories $\left(X_j^{(n)} \right)_{j=0, \dots, T}$, $n = 1, \dots, N$, we carry out the following procedure.

Step 1: At $i = T$ we set on each trajectory $\vartheta_T^{(n)} := \vartheta_T^{(n)}(M) := Z_T(X_T^{(n)})$ and $\left(M_j^{(n)} - M_T^{(n)} \right)_{T \leq j \leq T} = M_T^{(n)} - M_T^{(n)} = 0$ for $n = 1, \dots, N$.

Step 2: For $n = 1, \dots, N$ let $\left(M_j^{(n)} - M_{i+1}^{(n)} \right)_{i+1 \leq j \leq T}$ be constructed. For $i = T - 1$ down to $i = 0$, based on the N samples, we solve the regression problem

$$\left[\hat{\beta}^{(i)}, \hat{\gamma}^{(i)} \right] := \arg \min_{\beta \in \mathbb{R}^K, \gamma \in \mathbb{R}^{K_1}} \frac{1}{N} \sum_{n=1}^N \left(\vartheta_{i+1}^{(n)}(M) - \sum_{k=1}^K \beta_k \mathbf{m}_{i+1}^{(k,n)} - \sum_{k=1}^{K_1} \gamma_k \psi_k(i, X_i^{(n)}) \right)^2.$$

We then put

$$\begin{aligned}\widehat{\xi}_{i+1}^{(n)} &:= \sum_{k=1}^K \widehat{\beta}_k^{(i)} \mathbf{m}_{i+1}^{(k,n)}, \\ M_j^{(n)} - M_i^{(n)} &:= \left(M_j^{(n)} - M_{i+1}^{(n)} \right) + \widehat{\xi}_{i+1}^{(n)},\end{aligned}$$

and

$$\vartheta_i^{(n)}(M) := Z_i^{(n)} + \left(\vartheta_{i+1}^{(n)}(M) - \widehat{\xi}_{i+1}^{(n)} - Z_i^{(n)} \right)^+.$$

Step 3: We simulate \widetilde{N} new independent samples $\left(\widetilde{X}_j^{(n)} \right)_{j=0, \dots, T}$, $n = 1, \dots, \widetilde{N}$, which give rise to the new martingale samples

$$\widetilde{M}_i^{(n)} = \sum_{j=1}^i \sum_{k=1}^K \widehat{\beta}_k^{(j)} \widetilde{\mathbf{m}}_j^{(k,n)}, \quad k = 1, \dots, K, \quad n = 1, \dots, \widetilde{N}.$$

Then, an upper biased estimate for the upper bound is given by

$$\widehat{Y}_0^{\text{up}} := \frac{1}{\widetilde{N}} \sum_{n=1}^{\widetilde{N}} \max_{0 \leq i \leq T} \left(Z_i^{(n)}(\widetilde{X}_i^{(n)}) - \sum_{j=1}^i \sum_{k=1}^K \widehat{\beta}_k^{(j)} \widetilde{\mathbf{m}}_j^{(k,n)} \right). \quad (34)$$

Step 4: Based on the stopping rule

$$\tau_0(X_i) := \inf\{i \geq 0 : Z_i(X_i) \geq \sum_{k=1}^{K_1} \widehat{\gamma}_k^{(i)} \psi_k(i, X_i)\} \quad (35)$$

we put

$$\widehat{Y}_0^{\text{low}} := \frac{1}{\widetilde{N}} \sum_{n=1}^{\widetilde{N}} Z_{\tau_0(\widetilde{X}^{(n)})}^{(n)}(\widetilde{X}_{\tau_0(\widetilde{X}^{(n)})}^{(n)}).$$

which yields lower biased estimate to Y_0^* .

At this point, let us briefly compare our algorithm with the algorithm from Belomestny et. al. (2009). The methodology of Belomestny et. al. (2009) to compute dual martingales is built upon a procedure to numerically approximate Clark-Ocone derivatives of an approximative Snell envelope Y with respect to a Wiener filtration. The key ingredient in Belomestny et. al. (2009) is to approximate this Clark-Ocone derivative on a (fine) grid $\pi = \{t_0, \dots, t_N\}$ which contains the exercise grid $\{0, 1, \dots, T\}$ using the estimator,

$$Z_{t_j}^\pi := \frac{1}{\Delta_j^\pi} \mathbb{E}_{t_j}[\Delta^\pi W_j | Y_{i+1}], \quad (36)$$

where $\Delta_j^\pi = t_{j+1} - t_j$ and $\Delta^\pi W_j = W_{t_{j+1}} - W_{t_j}$. Due to (36), Belomestny et. al. (2009) morally requires to carry out a regression at each $t_j \in \pi$ on the fine grid π . However, our algorithm only needs to carry out regressions on the coarser grid of the possible exercise dates $\{0, \dots, T\} \subset \pi$. Moreover, note that (36) requires as an input some approximation of Y , which needs to be obtained by another method, such as the method of Longstaff and Schwartz (2001). Furthermore, if the grid π happens to be very fine, i.e. if $|\pi| = \sup_j |t_{j+1} - t_j| = \varepsilon$ for some very small $\varepsilon > 0$, the complexity increases, and also the right-hand of (36) becomes very large and even explodes as ε approaches zero. To circumvent these instabilities, Belomestny et. al. (2009) implement regressions on the coarser exercise time grid and then locally interpolate on the finer grid π . In our algorithm, these problems do not appear at all. Finally, we underline that obtaining numerically the Clark-Ocone derivative (36) in a non-Wiener filtration (e.g. filtrations generated by Lévy processes) is not so straightforward. In contrast, in our framework, the regression procedure (30) may include jump martingales as depicted in Remark 16.

7 Numerical examples

In this section we present the numerical results of the backward algorithm described in Section 6. The performance and accuracy of our algorithm is illustrated by testing it with two benchmark examples from the literature, a Bermudan basket-put on 5 assets and Bermudan max-call on 2 and 5 assets (see Bender et al. (2006a) and Belomestny et. al. (2009) respectively). In both examples, the risk-neutral dynamic of each asset is governed by

$$dX_t^d = (r - \delta)X_t^d dt + \sigma X_t^d dW_t^d, \quad d = 1, \dots, D,$$

where $D \in \mathbb{N}$ is the number of assets, W_t^d , $d = 1, \dots, D$, are independent one-dimensional Brownian motions, and r, δ and σ are constant real valued parameters. Exercise opportunities are equally spaced at times $T_j = \frac{jT}{J}$, $j = 0, \dots, J$. The discounted payoff from exercise at time t is given by

$$Z_t(X_t) = e^{-rt} \left(K - \frac{X_t^1 + \dots + X_t^D}{D} \right)^+ \quad \text{for the Bermudan basket-put,}$$

and

$$Z_t(X_t) = e^{-rt} (\max(X_t^1, \dots, X_t^D) - K)^+ \quad \text{for the Bermudan max-call,}$$

where we denote $X_t = (X_t^1, \dots, X_t^D)$. For both products, the time interval $[T_j, T_{j+1}]$, $j = 0, \dots, J-1$, is partitioned into L equally spaced subintervals of width $\Delta t = \frac{T}{N}$ with $N = J \times L$.

The implementation can be outlined as follows. We first simulate M independent samples of Brownian increments

$$\Delta W_i = (\Delta W_i^{1,(m)}, \dots, \Delta W_i^{D,(m)}), \quad i = 1, \dots, N, \quad m = 1, \dots, M.$$

Then the trajectories of $X_i^{(m)} = (X_i^{1,(m)}, \dots, X_i^{D,(m)})$, $i = 1, \dots, N$, $m = 1, \dots, M$, are given by

$$X_i^{d,(m)} = X_{i-1}^{d,(m)} \exp \left((r - \delta - \frac{1}{2}\sigma^2)\Delta t + \sigma \Delta W_i^{d,(m)} \right), \quad (37)$$

for $d = 1, \dots, D$ and initial data $X_0 = (X_0^1, \dots, X_0^D)$.

We now carry out the backward Monte Carlo regression algorithm as described in Section 6. In this Wiener setting, we recall Remark 17 (i) and choose as the spanning family of surely optimal martingales the Wiener integrals $\mathbf{m}_{i+1}^{(k)} = \int_{T_i}^{T_{i+1}} \varphi_k^c(s, X_s) dW_s$. More precisely, we solve in a first step the regression problem backward in time

$$\begin{aligned} (\hat{\beta}^{(i)}, \hat{\gamma}^{(i)}) := \arg \min_{(\beta, \gamma)} \frac{1}{M} \sum_{m=1}^M \left[\vartheta_{i+1}^{(m)} - \sum_{k=1}^K \beta_k \int_i^{i+1} \varphi_k(u, X_u^{(m)}) dW_u^{(m)} \right. \\ \left. - \sum_{k=1}^{K_1} \gamma_k \psi_k(i, X_i^{(m)}) \right]^2, \quad i = T-1, \dots, 0, \end{aligned} \quad (38)$$

for two families of basis functions $(\varphi_k) = (\varphi_k^{(d)})$ with $\varphi_k^{(d)} = \varphi_k^{(1)}$, and (ψ_k) , chosen as explained below. In (38) the Wiener integrals are approximated by the standard Euler scheme, using the same Brownian increments as in (37). Finally, a new independent simulation is launched and we estimate an upper bound \hat{Y}_0^{up} and a lower bound \hat{Y}_0^{low} by means of (34) and (34).

As one may expect, the choice of basis functions is crucial to obtain tight upper and lower bounds. In this respect, special information on the pricing problem may help us finding suitable basis functions. One way of retrieving additional information is to employ martingales representations and Malliavian calculus techniques to obtain more specific insights into the structure of the pricing dynamics. We illustrate this by considering the following stylized setting: By the Markov property of X , we have that $E_t(Z_T(X_T)) = f(t, X_t)$ for some measurable function $f(t, x)$ and $0 \leq t \leq T = T_J$. Let us assume that $f(t, x)$ is differentiable in x . Then, by Itô's formula and the fact that $E_t(Z_T)$ is a martingale we have

$$Z_T(X_T) - E_{T_{J-1}}(Z_T(X_T)) = \sum_{d=1}^D \sigma \int_{T_{J-1}}^T f_{x^d}(t, X_t) X_t^d dW_t^d.$$

Recall that $\hat{\vartheta}_T = Z_T$ and $E_{T_{J-1}}(Z_T(X_T))$ can be expressed in the following form

$$E_{T_{J-1}}(Z_T(X_T)) = e^{-rT_{J-1}} EP(T_{J-1}, X_{T_{J-1}}; T),$$

where $EP(t, x; T)$ is the price of the corresponding European option with maturity T at time t . Thus, it is natural to choose from time T to time T_{J-1}

European option values for the basis $(\psi_k(t, x))$ and the corresponding European deltas multiplied by the value of the underlying asset for the basis $(\varphi_k(t, x))$. Although for the following steps ($t < T_{J-1}$) there is no easy way to predict optimal choices of (ψ_k) and (φ_k) , the above analysis suggests to always include the still-alive European options into the basis (ψ_k) and include the information on the European deltas into the basis (φ_k) . In fact, based on similar arguments, this choice of basis functions were already proposed in Belomestny et. al. (2009).

7.1 Bermudan basket-put

In this example, we take the following parameter values,

$$r = 0.05, \quad \delta = 0, \quad \sigma = 0.2, \quad D = 5, \quad T = 3,$$

and

$$X_0^1 = \dots = X_0^D = x_0, \quad K = 100.$$

We perform the simulation of the underlying asset X from (37) with a time step size $\Delta t = 0.01$. For $T_j \leq t < T_{j+1}$, $j = 0, \dots, J-1$, we choose the set

$$\left\{ 1, Pol_3(X_t), Pol_3(EP(t, X_t; T_{j+1})), Pol_3(EP(t, X_t; T_J)) \right\}$$

as basis functions (ψ_k) , where $Pol_n(y)$ denotes the set of monomials of degree up to n in the components of a vector y and $EP(t, X; T)$ denotes the (approximated) value of a European basket-put with maturity T at time t . Recall that the family (ψ_k) serves as the regression basis for the continuation value. Further we choose

$$\left\{ 1, \left(X_t^d \frac{\partial EP(t, X_t; T_{j+1})}{\partial X_t^d} \right)_{1 \leq d \leq D}, \left(X_t^d \frac{\partial EP(t, X_t; T_J)}{\partial X_t^d} \right)_{1 \leq d \leq D} \right\}$$

as a regression basis (φ_k) spanning the family of the surely optimal martingales. Since there is no closed-form formula for the still-alive European basket-put, we use the moment-matching method to approximate their values (see e.g.

Brigo et al. (2004), and Lord (2006)). To this end, Let $S_t = \frac{X_t^1 + \dots + X_t^D}{D}$, and consider another asset G_t whose risk-neutral dynamic follows

$$dG_t = rG_t dt + \tilde{\sigma} G_t dW_t^1,$$

where $\tilde{\sigma}$ is a constant. The value of the European put on this asset can be easily computed by the well-known Black-Scholes formula, that is,

$$E[e^{-rT}(K - G_T)^+] = BS(G_0, r, \tilde{\sigma}, K, T). \quad (39)$$

If S_T and G_T have the same moments up to two, then the Black-Scholes price in (39) can be regarded as a good approximation for the value of the European

basket-put $E(e^{-rT}(K - S_T)^+)$, for details see Lord (2006). Since

$$E(S_T) = \frac{1}{D} \sum_{d=1}^D X_0^d e^{rT},$$

$$E(S_T^2) = \frac{1}{D^2} e^{2rT} \left(\sum_{i,j=1}^D X_0^i X_0^j \exp(1_{i=j} \sigma^2 T) \right)$$

and

$$E(G_T) = G_0 e^{rT}, \quad E(G_T^2) = G_0^2 e^{2rT + \tilde{\sigma}^2 T},$$

we can simply set

$$G_0 = \frac{1}{D} \sum_{d=1}^D X_0^d$$

and

$$\tilde{\sigma}^2 = \frac{1}{T} \ln \left(\frac{1}{(\sum_{d=1}^D X_0^d)^2} \sum_{i,j=1}^D X_0^i X_0^j \exp(1_{i=j} \sigma^2 T) \right).$$

The European deltas can be approximated by

$$\frac{\partial BS}{\partial G_0} \frac{\partial G_0}{\partial X_0^d} = -\mathcal{N}(-d_1) \frac{1}{D}, \quad d = 1, \dots, D,$$

where $d_1 = \frac{\ln(\frac{G_0}{K}) + (r + \frac{\tilde{\sigma}^2}{2})T}{\tilde{\sigma}\sqrt{T}}$ and \mathcal{N} denotes the cumulative standard normal distribution function. These formulas are straightforwardly extended to the pricing at times $t > 0$.

The numerical results are shown in Table 1. We use 1000 paths for estimating the surely optimal martingale and the continuation function via the regression procedure. Another 300000 paths are used to compute the lower bound and 100000 paths are used to compute the upper bound. Note that we have chosen a relatively small number of samples (1000) for estimating the martingale in the regression procedure. We do so because on the assumption that the family of uniformly integrable martingales is rich enough, the arguments leading to (24) yield that only a small number of samples are required for identifying a good approximation to a surely optimal martingale. We compare our results to the price intervals obtained in Bender et al. (2006a) which are displayed in the last column of Table 1. In our C++ implementation, the run-times for computing one set of lower and upper bounds are in the range of 15-20 minutes.

7.2 Bermudan max-call

We use the same parameter values as in Section 7.1 except $\delta = 0.1$ and $D = 2$ or 5. As in the previous example we use European (call) options in the basis

Table 1: Lower and upper bounds for Bermudan basket-put on 5 assets with parameters $r = 0.05$, $\delta = 0$, $\sigma = 0.2$, $K = 100$, $T = 3$ and different J and x_0

J	x_0	Low (SE)	Up (SE)	BKS Price Interval
3	90	10.000 (0.000)	10.000 (0.000)	[10.000, 10.004]
	100	2.164 (0.007)	2.172 (0.001)	[2.154, 2.164]
	110	0.539 (0.004)	0.551 (0.001)	[0.535, 0.540]
6	90	10.000 (0.000)	10.000 (0.000)	[10.000, 10.000]
	100	2.407 (0.006)	2.432 (0.001)	[2.359, 2.412]
	110	0.573 (0.003)	0.609 (0.001)	[0.569, 0.580]
9	90	10.000 (0.0000)	10.008 (0.0003)	[10.000, 10.005]
	100	2.475 (0.0063)	2.522 (0.0013)	[2.385, 2.502]
	110	0.5915 (0.0034)	0.6353 (0.0009)	[0.577, 0.600]

(ψ_k) and the corresponding deltas in the basis (φ_k) . The value of the European max-call option is computed by the following formula (Johnson (1987)),

$$C_{max} = \sum_{l=1}^D X_0^l \frac{e^{-\delta T}}{\sqrt{2\pi}} \int_{(-\infty, d_+^l]} \exp[-\frac{1}{2}z^2] \prod_{\substack{l'=1 \\ l' \neq l}}^D \mathcal{N}\left(\frac{\ln \frac{X_0^l}{X_0^{l'}}}{\sigma\sqrt{T}} - z + \sigma\sqrt{T}\right) dz \\ - K e^{-rT} + K e^{-rT} \prod_{l=1}^D (1 - \mathcal{N}(d_-^l)), \quad (40)$$

where

$$d_-^l := \frac{\ln \frac{X_0^l}{K} + (r - \delta - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}, \quad d_+^l = d_-^l + \sigma\sqrt{T}.$$

Moreover, straightforward computations reveal that the deltas are given by

$$\frac{\partial C_{max}}{\partial X_0^l} = \frac{e^{-\delta T}}{\sqrt{2\pi}} \int_{(-\infty, d_+^l]} \exp[-\frac{1}{2}z^2] \prod_{\substack{l'=1 \\ l' \neq l}}^D \mathcal{N}\left(\frac{\ln \frac{X_0^l}{X_0^{l'}}}{\sigma\sqrt{T}} - z + \sigma\sqrt{T}\right) dz, \quad (41)$$

and that C_{max} satisfies the linear homogeneity³

$$C_{max} = \sum_{l=1}^D X_0^l \frac{\partial C_{max}}{\partial X_0^l} + K \frac{\partial C_{max}}{\partial K}. \quad (42)$$

The numerical results are shown in Table 2. They are based on 1000 paths for the regression procedure, 300000 paths for computing the lower bound and 100000 paths for computing the upper bound. As before, we have chosen a relatively small number of samples (1000) for estimating the martingale in the

³Compare also with (Johnson, 1987, eq. (9)).

Table 2: Lower and upper bounds for Bermudan max-call with parameters $r = 0.05$, $\delta = 0.1$, $\sigma = 0.2$, $K = 100$, $T = 3$ and different D and x_0 .

D	x_0	Low (SE)	Up (SE)	A&B price interval
2	90	8.0556 (0.0219)	8.15655 (0.0034)	[8.053, 8.082]
	100	13.8850 (0.0276)	14.0293 (0.0044)	[13.892, 13.934]
	110	21.3671 (0.0319)	21.5319 (0.0048)	[21.316, 21.359]
5	90	16.5973 (0.0296)	16.7963 (0.0058)	[16.602, 16.655]
	100	26.1325 (0.0356)	26.3803 (0.0072)	[26.109, 26.292]
	110	36.7348 (0.0403)	37.0856 (0.0082)	[36.704, 36.832]

regression procedure. This is again allowed because the arguments leading to (24) and the assumption that the choice of the basis functions indeed equips us with a rich enough family of uniformly integrable martingales yield that only a small number of samples are required for identifying a good approximation to a surely optimal martingale. The integral expressions from (40) and (41) are numerically evaluated using a simple adaptive Gauss-Kronrod procedure with 31 points. The price intervals in the last column are quoted from Andersen and Broadie (2004). In our C++ implementation, for each set of lower and upper bounds, we observe run-times that are in the range of 10-25 minutes, with the longer computation times for the 5-dimensional case.

Concluding remark

The numerical results presented in Tables 1, 2 due to our new algorithm may be considered as very satisfactory given the decreased computation times (which are in the order of minutes in a C++ compiled implementation). In this respect it should be noted that computing upper bounds (in a rather generic way) in order of minutes is a considerable improvement compared to Bender et al. (2006a), whose upper bounds are computed with nested Monte Carlo simulation requiring higher computation time, and of comparable range to Belomestny et. al. (2009). Moreover, the algorithm delivers fast and surprisingly good lower bounds while the upper bounds are about the same range as the ones obtained with the algorithm in Belomestny et. al. (2009). Needless to say that, as for the method of Belomestny et. al. (2009), the performance of the here presented algorithm will highly depend on the choice of the basis functions. An in depth treatment of this issue is considered beyond scope however.

8 Appendix

We present in this section some well-known facts from theory of empirical processes which are used to establish the relation (23) in Section 6.1.

Let $\{\vartheta^q : q \in Q\}$ be a family of random variables and let for each $q \in Q$, $\vartheta^{q,1}, \dots, \vartheta^{q,N}$ be i.i.d. samples of ϑ^q . For each $q \in Q$ we consider the unbiased

variance estimator

$$\text{Var}^{(N)} \vartheta^q := \frac{1}{N-1} \sum_{n=1}^N \left(\vartheta^{q,n} - \overline{\vartheta_N^q} \right)^2 \quad \text{with} \quad \overline{\vartheta_N^q} := \frac{1}{N} \sum_{n=1}^N \vartheta^{q,n},$$

hence $E \text{Var}^{(N)} \vartheta^q = \text{Var} \vartheta^q$ for $q \in Q$. From standard statistical theory it is well known that with $\mu^q := E \vartheta^q$

$$\text{Var} \left(\text{Var}^{(N)} \vartheta^q \right) \leq \frac{1}{N} E (\vartheta^q - \mu^q)^4 = \frac{(\text{Var} \vartheta^q)^2}{N} E \left(\frac{\vartheta^q - \mu^q}{\sqrt{\text{Var} \vartheta^q}} \right)^4.$$

Now, as a mild condition we assume that,

$$E \left(\frac{\vartheta^q - \mu^q}{\sqrt{\text{Var} \vartheta^q}} \right)^4 \leq C \quad \text{for all } q \in Q.$$

For example, this holds if $E \exp [\lambda |\vartheta^q|] < \infty$ for some $\lambda > 0$ and all $q \in Q$. We so have in particular

$$\text{Var} \left(\text{Var}^{(N)} \vartheta^{q_N^\circ} \right) \leq \frac{C}{N} \left(\text{Var} \vartheta^{q_N^\circ} \right)^2 \quad \text{and} \quad \text{Var} \left(\text{Var}^{(N)} \vartheta^{q^\circ} \right) \leq \frac{C}{N} \left(\text{Var} \vartheta^{q^\circ} \right)^2,$$

while, strictly speaking, the randomness of $\vartheta^{q_N^\circ}$ in the first inequality is ignored. However, by considering a next from $\vartheta^{q_N^\circ}$ independent sample, we can show that this is not really essential (the details would go beyond the scope of the optional analysis of Section 6.1 and are therefore omitted). From standard empirical probability theory it now follows that for any (small) $0 \leq \alpha \ll 1$, there is a suitable quantile coefficient c_α (particularly not depending on N) such that

$$\begin{aligned} \mathbb{P} \left[\text{Var}^{(N)} \vartheta^q \leq \text{Var} \vartheta^q \left(1 + c_\alpha \sqrt{\frac{C}{N}} \right), \text{Var} \vartheta^q \leq \text{Var}^{(N)} \vartheta^q + c_\alpha \sqrt{\frac{C}{N}} \text{Var} \vartheta^q \right] \\ \geq 1 - \alpha, \end{aligned}$$

for $q \in \{q_N^\circ, q^\circ\}$ which implies (23).

Acknowledgements

J.S. is grateful to Denis Belomestny for interesting discussions on this topic, and to Christian Bender for an inspiring remark, which led to counterexample Example 14.

References

- L. Andersen, M. Broadie (2004). A primal-dual simulation algorithm for pricing multidimensional American options. *Management Sciences*, **50**, No. 9, 1222-1234.

- V. Bally, G. Pages (2003). A quantization algorithm for solving multidimensional discrete optimal stopping problem. *Bernoulli*, **9**(6), 1003-1049.
- D. Belomestny, C. Bender, and J. Schoenmakers (2009). True upper bounds for Bermudan products via non-nested Monte Carlo. *Mathematical Finance*, **19**, No. 1, 53–71.
- C. Bender, A. Kolodko, J. Schoenmakers (2008). Enhanced policy iteration for American options via scenario selection. *Quant. Finance* **8**, No. 2, 135–146.
- C. Bender, A. Kolodko, J. Schoenmakers (2006). Iterating cancellable snowballs and related exotics. *RISK*, September 2006 pp. 126–130.
- C. Bender, A. Kolodko, J. Schoenmakers (2006a). Policy iteration for american options: overview. *Monte Carlo Methods and Appl.*, **12**, No. 5-6, pp. 347–362.
- D. Brigo, F. Mercurio, F. Rapisarda, and R. Scotti (2004). Approximated moment-matching dynamics for basket-options pricing. *Quantitative Finance*, **4**, 1–16.
- M. Broadie, P. Glasserman (2004). A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, **7**(4), 35–72.
- D. Brown, J. Smith, P. Sun (2009). Information Relaxations and Duality in Stochastic Dynamic Programs. *Operations Research*, **58**(4), 785–801.
- J. Carriere (1996). Valuation of early-exercise price of options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, **19**, 19-30.
- N. Chen and P. Glasserman (2007). Additive and Multiplicative Duals for American Option Pricing. *Finance and Stochastics*, **11**, 153–179.
- V. Desai, V. Farias, C. Moallemi (2010). Pathwise Optimization for Optimal Stopping Problems. Working paper.
- P. Glasserman (2003). *Monte Carlo Methods in Financial Engineering*. Springer.
- M. Haugh, L. Kogan (2004). Pricing American options: a duality approach. *Operations Research*, **52**, No. 2, 258–270.
- F. Jamshidian (2007). The duality of optimal exercise and domineering claims: A Doob-Meyer decomposition approach to the Snell envelope. *Stochastics*, **79**, No. 1–2, 27–60.
- H. Johnson (1987). Options on the maximum or the minimum of several assets. *Journal of Financial and Quantitative Analysis* **22**, 227–83.
- P.E. Kloeden, E. Platen (1992). *Numerical solution of stochastic differential equations*. Springer Verlag Berlin.

- A. Kolodko, J. Schoenmakers (2004). Upper bounds for Bermudan style derivatives. *Monte Carlo Methods and Appl.*, **10**, No. 3–4, 331–343.
- A. Kolodko, J. Schoenmakers (2006). Iterative construction of the optimal Bermudan stopping time. *Finance and Stochastics*, **10**, 27–49.
- H. Liero (1989). Strong Uniform Consistency of Nonparametric Regression Function Estimates. *Probab. Th. Rel. Fields*, **82**, 587–614.
- F.A. Longstaff, E.S. Schwartz (2001). Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, **14**, 113–147.
- R. Lord (2005). Partially exact and bounded approximations for arithmetic Asian options. *Journal of Computational Finance*, **10**, no. 2, pp. 1–52.
- J. Neveu (1975). *Discrete Parameter Martingales*. North-Holland, Amsterdam.
- L.C.G. Rogers (2002). Monte Carlo valuation of American options. *Mathematical Finance*, **12**, 271–286.
- L.C.G. Rogers (2010). Dual Valuation and Hedging of Bermudan Options. *SIAM J. Financial Math.*, **1**, 604–608.
- J. Schoenmakers (2005). *Robust Libor Modelling and Pricing of Derivative Products*. Chapman & Hall/CRC.
- J. Schoenmakers, J. Huang (2010). Optimal dual martingales and their stability; fast evaluation of Bermudan products via dual backward regression. *WIAS Preprint 1574*.
- J. Tsitsiklis, B. Van Roy (2001). Regression methods for pricing complex American style options. *IEEE Trans. Neural. Net.*, **12**, 694–703.